

NovaSeq™ X シリーズによる 高精度の次世代 シーケンス

XLEAP-SBS™ケミストリーが、
ハイスループットのゲノミクス
アプリケーションに対応する
優れたデータ品質を提供

- 全ゲノムシーケンス
- 全エクソームシーケンス
- 全トランスクリプトームシーケンス
- 全ゲノムメチル化シーケンス
- シングルセルマルチオームシーケンス

illumina®

はじめに

NovaSeq XおよびNovaSeq X Plusシステムに搭載されたテクノロジーの進歩により、スループットと生産性の向上が実現し、生産規模シーケンスの経済性に革新をもたらします。NovaSeq Xシリーズは、先進的なケミストリー、超高解像度光学系、統合型の二次解析、および操作の簡便性を組み合わせた、これまでで最もパワフルでコスト効率の高いシーケンスシステムです。

NovaSeq Xシリーズは、実績のあるイルミナ Sequence by Synthesis (SBS) ケミストリーを、より高速で高精度、よりロバストに進化させたXLEAP-SBSケミストリーを採用しています。XLEAP-SBS試薬は、期待される高品質のデータを犠牲にすることなくスループットを最大化するために、性能と速度が最適化されています。

このアプリケーションノートでは、NovaSeq Xシリーズが、全ゲノムシーケンス、全エクソームシーケンス、全トランスクリプトームシーケンス、メチル化シーケンス、シングルセルマルチオミクスなどの主要なメソッドに対して、NovaSeq 6000システムと同等またはそれを超えるデータ品質を提供することを示します。

方法

全ゲノムシーケンス

全ゲノムライブラリーは、TruSeq™ PCR-Free Prepキット（イルミナ、カタログ番号：20015963）を使用し、NA12878ゲノムDNA（gDNA）（Coriell Institute for Medical Research）から調製しました。

シーケンスは、NovaSeq X Series 10B Reagent Kit (300 cycles)（イルミナ、カタログ番号：20085594）を使用し、NovaSeq X Plusシステムで、151 bp × 2ラン構成（複数回のランで42サンプル）を使用して実施しました。また、比較のために、同じライブラリーをNovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles)（イルミナ、カタログ番号：20028312）を使用し、NovaSeq 6000システムで151 bp × 2ラン構成（1回のランで24サンプル）を使用してシーケンスしました。

二次データ解析は、DRAGEN™ Germlineパイプラインv4.1のクラウドベースワークフローを使って実施しました。NovaSeq X PlusシステムとNovaSeq 6000システム間のバリエーションコール性能を比較するために、シーケンスデータを30xカバレッジにダウンサンプリングしました。

エクソームシーケンス

エクソームライブラリーは、Illumina DNA Prep with Enrichment, (S) Tagmentation（イルミナ、カタログ番号：2002554）を使用してNA12878ゲノムDNA（gDNA）（Coriell Institute for Medical Research）から調製し、Twist Comprehensive Exome Panel（Twist Bioscience、カタログ番号：102033）によってターゲットゲノム領域をキャプチャーしました。

シーケンスは、NovaSeq X Series 10B Reagent Kit (300 cycles)を使用し、NovaSeq X Plusシステムで、101 bp × 2ラン構成（複数回のランで753サンプル）を使用して実施しました。また、比較のために、同じライブラリーをNovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles)を使用し、NovaSeq 6000システムで101 bp × 2ラン構成（単一レーンで48サンプル）でシーケンスしました。

二次データ解析は、DRAGEN Enrichmentパイプラインv4.0.3のクラウドベースワークフローを使って実施しました。バリエーションコールの精度は、Genome In A Bottle (GIAB) v3.3.2 Truthセットおよびhg19-alt-awareリファレンスゲノムに対して評価しました。^{1,2} NovaSeq X PlusシステムとNovaSeq 6000システム間のバリエーションコール性能を比較するために、シーケンスデータはサンプルあたり3,000万リードペアにダウンサンプリングしました。

全トランスクリプトームシーケンス

トータルRNAおよびメッセンジャーRNA（mRNA）ライブラリーは、白血病細胞株RNAのHL-60（Thermo Fisher Scientific、カタログ番号：AM7836）およびK562（BioChain、カタログ番号：R1255820-50）、乳がん細胞株RNAのMCF7（BioChain、カタログ番号：R1255830-50）から、Illumina Stranded Total RNA Prep with Ribo-Zero Plus（イルミナ、カタログ番号：20040529）およびIllumina Stranded mRNA Prep（イルミナ、カタログ番号：20040534）を使用して調製しました。

シーケンスは、NovaSeq X Series 10B Reagent Kit (300 cycles)を使用してNovaSeq X Plusシステムで実施し、75 bp × 2ラン構成で、ライブラリー調製由来の1塩基目のTオーバーハングを回避するためにカスタムダークサイクルレシピを使用しました。³シーケンスは、トータルRNA-Seq573サンプル、mRNA-Seq2,304サンプルに対し複数回のランを行いました。また、比較のために、同じライブラリーをNovaSeq 6000 S2 Reagent Kit v1.5 (200 cycles) (イルミナ、カタログ番号: 20028315) を使用し、76 bp × 2ラン構成でNovaSeq 6000システムでシーケンスしました。トータルRNA-SeqとmRNA-Seqの両方について、それぞれ96サンプルを単一レーンでランを実施しました。

二次データ解析は、DRAGEN RNA Pipeline v4.1のクラウドベースワークフローを使って実施しました。遺伝子発現データを比較するために、シーケンスデータはすべてのサンプルについて1,000万リードにダウンサンプリングしました。データは、Genome Reference Consortium Human GRCh38 (h38 アセンブリ) に対してアライメントしました。²

全ゲノムメチル化シーケンス

メチル化ライブラリーは、Illumina DNA Prepライブラリー調製キット (Illumina、カタログ番号: 20060059) と共に、Zymo-Seq WGBS Library Kit (Zymo Research、カタログ番号: D5465) を使用し、マッチするヒトの脳と脾臓のサンプルセット (Zymo Research、カタログ番号: D5018) のレプリケート (NovaSeq X Plusシステムではそれぞれ8レプリケート、NovaSeq 6000システムではそれぞれ5レプリケート) から調製しました。⁴ 非メチル化大腸菌コントロールは、シトシン変換効率 (99.5%より高くなると推測) を得るために、0.25%で添加しました。

シーケンスは、NovaSeq X Series 10B Reagent Kit (300 cycles)を使用し、NovaSeq X Plusシステムで、151 bp × 2ラン構成 (1回のランで16サンプル) を使用して実施しました。比較のために、ライブラリーをNovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles)を使用し、NovaSeq 6000システムで151 bp × 2ラン構成 (1回のランで10サンプル) でシーケンスしました。

メチル化解析は、DRAGEN Methylationパイプラインのクラウドベースワークフローを使って実施しました。シーケンスデータは下流の解析のために、サンプルあたり5億リードにダウンサンプリングしました。

シングルセルマルチオミクス

Chromium Single Cell Multiome ATAC + Gene Expressionは、シングルセル解像度で遺伝子発現とエピジェネティックシグネチャーを統合したリードアウトを提供します。シングルセルRNAシーケンス (scRNA-Seq) およびsingle-cell assay for transposase accessible chromatin (scATAC-Seq) 用のサンプルは、AllCellsから取得した、健康な男性ドナー (35歳未満) の凍結保存されたヒト末梢血単核球 (PBMC) から一緒に調製しました。核は、10x GenomicsのDemonstrated Protocol『Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing』(CG000365 Rev A) の記載に従って分離しました。ペアのscRNA-SeqライブラリーとscATAC-Seqライブラリーは、『Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide』(CG000338 Rev B) の記載に従って分離した核から作製しました。

シーケンスは、NovaSeq X Series 10B Reagent Kit (300 cycles)を使用し、NovaSeq X Plusシステムで (1回のランで80レプリケート) 実施しました。また、比較のために、同じライブラリーをNovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles)を使用し、NovaSeq 6000システムでシーケンスしました (1回のランで10レプリケート)。ラン構成は、10x Genomicsによって提供された次のパラメーターに従ってセットアップしました。(マルチオームscRNA-Seqライブラリーには、リード1は28サイクル、i7とi5インデックスリード10サイクル、リード2は90サイクル。マルチオームscATAC-Seqライブラリーには、リード1は50サイクル、i7インデックスリード8サイクル、i5インデックスリード24サイクル、リード2は49サイクル。)

データ解析は、Cell Ranger ARC解析パイプラインv2.0 (10x Genomics) を使用して実施し、シングルセル内の転写産物とクロマチンアクセシビリティピークを計測しました。

結果

NovaSeq X Plusシステムは、NovaSeq 6000システムと比べてスループットを大幅に向上させることができます。NovaSeq X 10BフローセルとNovaSeq 6000 S4フローセルはどちらもフローセルあたり150 bp × 2のシーケンスデータから最大3 Tbを出力できます。ただし、NovaSeq Xシリーズのランタイムは、NovaSeq 6000システムのランタイムのほぼ半分です (表1)。

表1: 大幅に短いランタイムで同等のシーケンス出力を実現

メトリクス	NovaSeq 6000 S4フローセル	NovaSeq X Plus 10Bフローセル
ランあたり100 bp × 2の出力	1.6~4 TB	~2~4 Tb
2 × 100 bpのランタイム	~36時間	~22時間
ランあたり150 bp × 2の出力	2.4~6 Tb	~3~6 Tb
2 × 150 bpのランタイム	~44時間	~24時間

全ゲノムシーケンス

1塩基変異 (SNV) および挿入と欠失 (Indel) の両方の精度とコール率を含む全ゲノムシーケンス (WGS) 解析メトリクスを評価しました。NovaSeq X PlusシステムとNovaSeq 6000システムはどちらも、高品質のデータと高精度のバリエーションコールを提供しました (表2、表3)。これらのデータは、NovaSeq XシリーズでのWGSの結果がNovaSeq 6000システムの性能と同等か、それを上回っていることを示しています。

表2: WGSのシーケンスランメトリクス

メトリクス	NovaSeq 6000 システム	NovaSeq X Plus システム
ラン構成	151 bp × 2	151 bp × 2
Q30以上のリード1の塩基	92.17%	95.89%
Q30以上のリード2の塩基	89.60%	94.30%
リード1のエラー率	0.25%	0.15%
リード2のエラー率	0.24%	0.23%

単一のフローセルランからのメトリクスは、レーン数が異なる複数のフローセル全体を平均化したものです。すべてのランは出力に関して公表されている仕様を満たしました。レーンあたりの出力は、NovaSeq 6000 S4フローセルとNovaSeq X 10Bフローセルの間で同等ではありません。

表3: WGSの二次解析メトリクス

メトリクス	NovaSeq 6000 システム	NovaSeq X Plus システム
深度	30×	30×
常染色体カバレッジ	31×	31×
SNP総数	3,041,268	3,041,454
Het:Hom比	1.59	1.60
Ti:Tv比	1.99	1.98
SNP精度	99.95%	99.95%
SNPコール率	99.95%	99.96%
Indel精度	99.64%	99.60%
Indelコール率	99.61%	99.57%
リード1ミスマッチ塩基	0.48%	0.36%
リード2ミスマッチ塩基	0.61%	0.43%
平均されたサンプル数	24	42

全エクソームシーケンス

評価した全エクソームシーケンス (WES) の一次解析および二次解析メトリクスには、クオリティスコア、エラー率、常染色体コーラビリティ、アライメントされたリード割合、カバレッジの均一性、SNVとIndel両方の精度とコール率が含まれました。NovaSeq X PlusシステムとNovaSeq 6000システムはどちらも、高品質のデータと高精度のバリエーションコールを提供しました (表4、表5)。これらのデータは、NovaSeq XシリーズのWES結果がNovaSeq 6000システムの性能と同等であることを示しています。

表4: WESのシーケンスランメトリクス^a

メトリクス	NovaSeq 6000システム	NovaSeq X Plusシステム
ラン構成	101 bp × 2	101 bp × 2
Q30以上のリード1の塩基	92.06%	96.63%
Q30以上のリード2の塩基	90.96%	96.24%
リード1のエラー率	N/A ^b	0.10%
リード2のエラー率	N/A ^b	0.21%

- a. 単一のフローセルランからのメトリクスは、レーン数が異なる複数のフローセル全体を平均化したものです。すべてのランは出力に関して公表されている仕様を満たしました。レーンあたりの出力は、NovaSeq 6000 S4フローセルとNovaSeq X 10Bフローセルの間で同等ではありません。
- b. これらのNovaSeq 6000システムのランでPhiXコントロールを使用しなかったため、エラー率は示されていません。

表5: WESの二次解析メトリクス

メトリクス	NovaSeq 6000システム	NovaSeq X Plusシステム
常染色体コーラビリティ	97.49%	97.53%
アライメントリード	99.28%	99.11%
カバレッジ均一性	97.17%	97.29%
SNP精度	99.77%	99.77%
SNPコール率	98.20%	98.30%
Indel精度	97.57%	97.36%
Indelコール率	88.53%	89.05%
平均されたサンプル数	48	753

全トランスクリプトームシーケンス

NovaSeq X PlusシステムとNovaSeq 6000システムはどちらも、トータルRNA-Seq (表6) とmRNA-Seq (表7) のデータ品質について、公表されている仕様を上回りました。転写物の定量化より、トータルRNA-Seq (図1) とmRNA-Seq (図2) に対する2つのプラットフォーム間の優れた一致が示されました ($R^2 > 0.99$)。これらのデータは、NovaSeq Xシリーズの全トランスクリプトームシーケンスがNovaSeq 6000システムの性能と同等か、または上回るデータ品質を産生することを示しています。

表6: トータルRNA-Seqのシーケンスランメトリクス

メトリクス	NovaSeq 6000システム	NovaSeq X Plusシステム
ラン構成	76 bp × 2	75 bp × 2
Q30以上のリード1の塩基	91.83%	96.82%
Q30以上のリード2の塩基	90.52%	96.37%
リード1のエラー率	0.44%	0.07%
リード2のエラー率	1.17%	0.15%
平均されたサンプル数	96	573

単一のフローセルランからのメトリクスは、レーン数が異なる複数のフローセル全体を平均化したものです。すべてのランは出力に関して公表されている仕様を満たしました。レーンあたりの出力は、NovaSeq 6000 S4フローセルとNovaSeq X 10Bフローセルの間で同等ではありません。

表7: mRNA-Seqのシーケンスランメトリクス

メトリクス	NovaSeq 6000システム	NovaSeq X Plusシステム
ラン構成	76 bp × 2	75 bp × 2
Q30以上のリード1の塩基	91.47%	96.03%
Q30以上のリード2の塩基	89.92%	95.65%
リード1のエラー率	0.74%	0.09%
リード2のエラー率	1.32%	0.16%
平均されたサンプル数	96	2,304

単一のフローセルランからのメトリクスは、レーン数が異なる複数のフローセル全体を平均化したものです。すべてのランは出力に関して公表されている仕様を満たしました。レーンあたりの出力は、NovaSeq 6000 S4フローセルとNovaSeq X 10Bフローセルの間で同等ではありません。

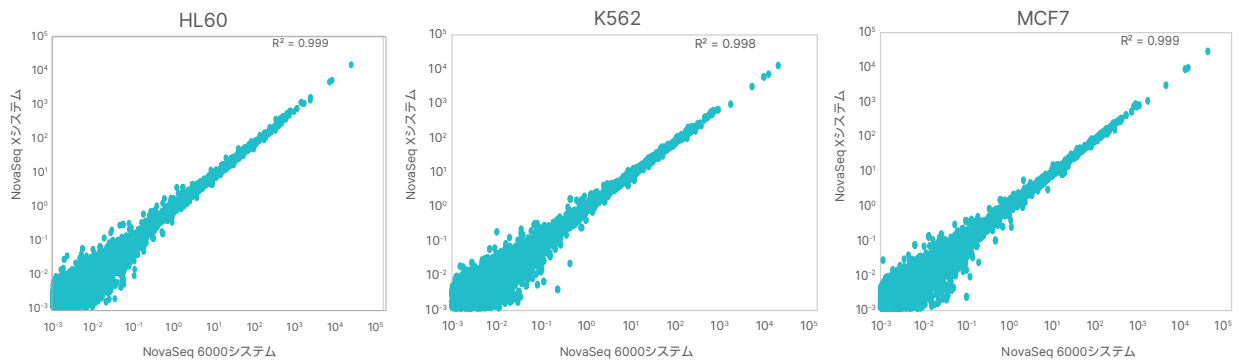


図1:全トランスクリプトームのトータルRNA-Seqの相関:がん細胞株 (HL-60、K562、MCF7) でのTranscripts Per Million値 (TPM)。

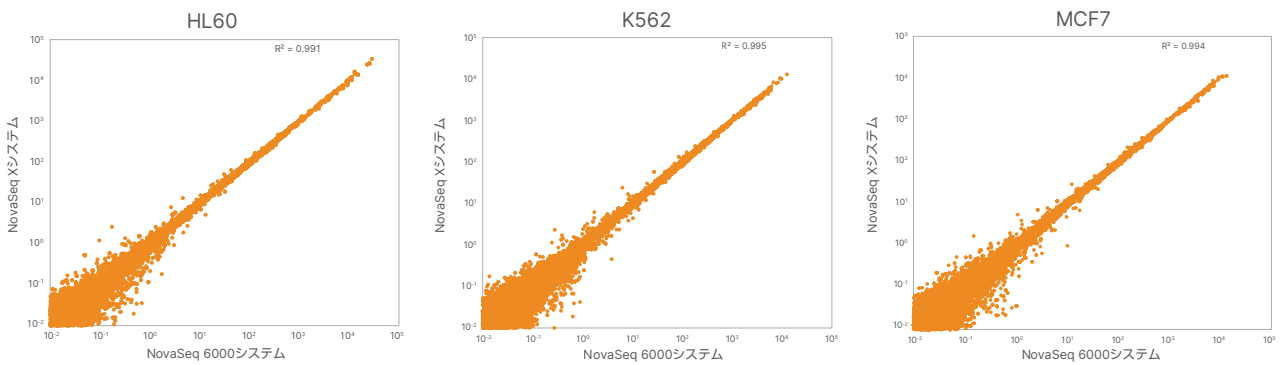


図2:全トランスクリプトームのmRNA-Seqの相関:がん細胞株 (HL-60、K562、MCF7) でのTranscripts Per Million値 (TPM)。

メチル化シーケンス

全ゲノムメチル化メトリクスを評価して、NovaSeq XシリーズとNovaSeq 6000システムの性能を比較しました。NovaSeq X PlusシステムとNovaSeq 6000システムではどちらも、メチル化シトシン率を定量する数値が製品文書から期待される値と一致しました。(図3A)。一致するライブラリーについては、NovaSeq X Plusシステムでより高いマッピング効率が検出されました(図3B)。全ゲノムカバレッジ分布プロットでは、NovaSeq X PlusシステムとNovaSeq 6000システム間で高カバレッジのCpG (> 50x)の増加が認められ、同等の結果が示されました。(図4)。

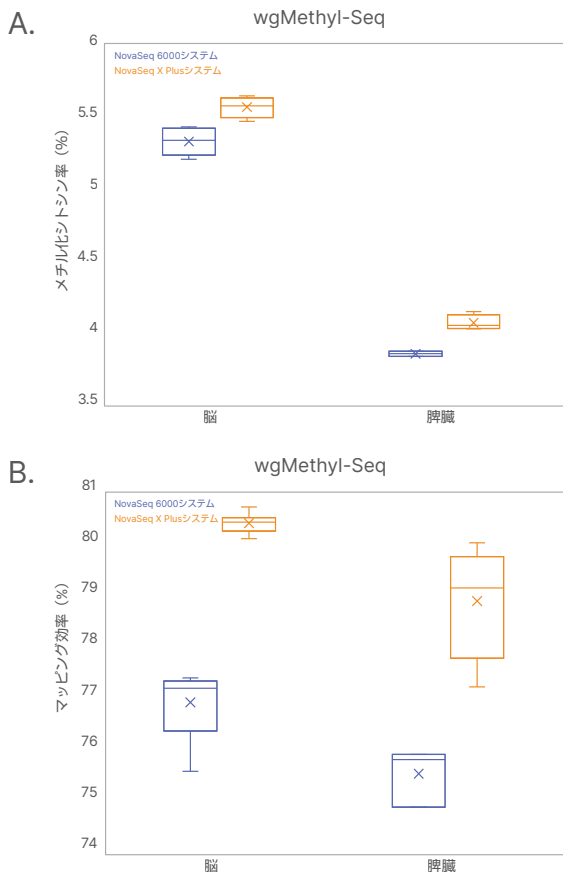


図3: 全ゲノムメチル化シーケンス: Zymo-Seq WGBS結果。NovaSeq X PlusシステムおよびNovaSeq 6000システムの (A) メチル化シトシン率および (B) マッピング効率を示しています

バイサルファイト変換または酵素変換により、ライブラリー調製中に非メチル化シトシンがウラシルに変化します。その結果、ライブラリーのバランスが崩れるため、これまでシーケンスが困難であり、塩基の多様性を高めるために通常は高い割合 (> 5%) のPhiXが必要でした。NovaSeq Xシリーズでは、低い割合 (1%) のPhiXでも十分高品質のランを達成することができました (表8)。

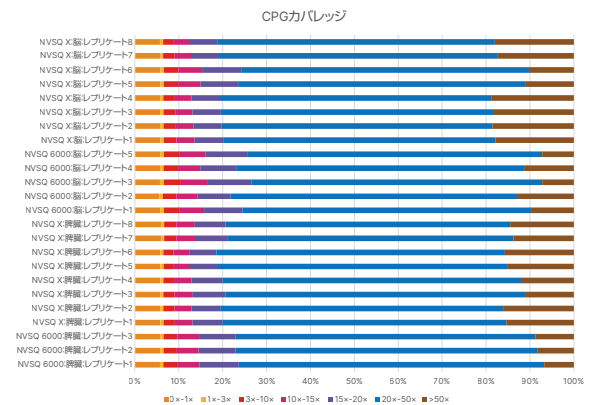


図4: 全ゲノムメチル化シーケンスのゲノムカバレッジ: Zymo-Seq WGBS結果。NovaSeq X PlusシステムおよびNovaSeq 6000システムのCpGカバレッジ分布を示しています。

表8: メチル化シーケンスのシーケンスランメトリクス

メトリクス	NovaSeq 6000システム	NovaSeq X Plusシステム
ラン構成	151 bp × 2	151 bp × 2
Q30以上のリード1の塩基	89.01%	91.95%
Q30以上のリード2の塩基	86.75%	93.09%
リード1のエラー率	0.30%	0.14%
リード2のエラー率	0.59%	0.25%
平均されたサンプル数	10	16

単一のフローセルランからのメトリクスは、レーン数が異なる複数のフローセル全体を平均化したものです。すべてのランは出力に関して公表されている仕様を満たしました。レーンあたりの出力は、NovaSeq 6000 S4フローセルとNovaSeq X 10Bフローセルの間で同等ではありません。

シングルセルマルチオミクス

遺伝子発現を測定するためのscRNA-Seqおよびクロマチンアクセシビリティを測定するためのscATAC-Seqを含む、シングルセルマルチオミクスアッセイの性能メトリクスを評価しました。NovaSeq X PlusシステムとNovaSeq 6000システムはデータ品質に対して公表された仕様を上回りました(表9、表10)。scRNA-Seq遺伝子発現(表5)とscATAC-Seqクロマチンアクセシビリティ(表6)のt-SNEプロットより、NovaSeq X PlusシステムとNovaSeq 6000システム間の優れた一致が示されました。

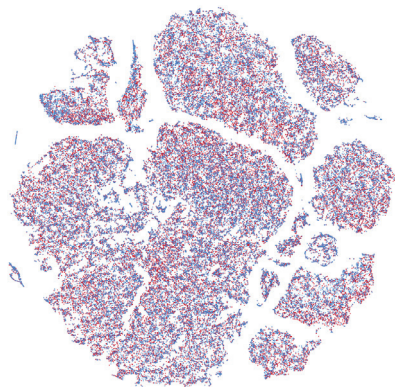


図5: シングルセルマルチオーム遺伝子発現: NovaSeq X Plusシステム(青)およびNovaSeq 6000システム(赤)でシーケンスしたscRNA-Seqライブラリーのt-SNEプロット。

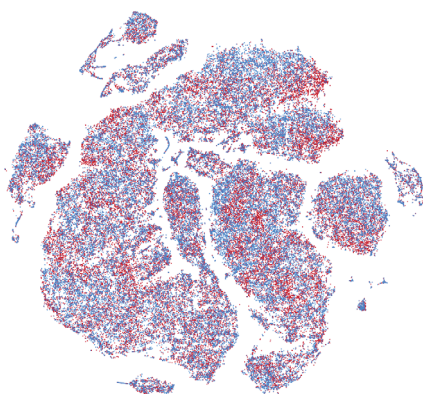


図6: シングルセルマルチオームクロマチンアクセシビリティ: NovaSeq X Plusシステム(青)およびNovaSeq 6000システム(赤)でシーケンスしたscATAC-Seqライブラリーのt-SNEプロット。

表9: シングルセルマルチオームscRNA-Seqのシーケンスランメトリクス

メトリクス	NovaSeq 6000システム	NovaSeq X Plusシステム
ラン構成		
リード1	28 bp	28 bp
インデックス1	10 bp	10 bp
インデックス2	10 bp	10 bp
リード2	90 bp	90 bp
Q30以上のリード1の塩基	97.11%	97.35%
Q30以上のリード2の塩基	95.01%	95.93%
リード1のエラー率	0.04%	0.04%
リード2のエラー率	0.17%	0.15%
平均されたサンプル数	10	80
検出された遺伝子数	25,975	25,743
細胞あたりのUMIカウント中央値	2,790	2,571
サンプルあたりの推測される細胞数	3,880	3,882

単一のフローセルランからのメトリクスは、レーン数が異なる複数のフローセル全体を平均化したものです。すべてのランは出力に関して公表されている仕様を満たしました。レーンあたりの出力は、NovaSeq 6000 S4フローセルとNovaSeq X 10Bフローセルの間で同等ではありません。

表10: シングルセルマルチオームscATAC-Seqのシーケンスランメトリクス

メトリクス	NovaSeq 6000システム	NovaSeq X Plusシステム
ラン構成		
リード1	50 bp	50 bp
インデックス1	8 bp	8 bp
インデックス2	24 bp	24 bp
リード2	49 bp	49 bp
Q30以上のリード1の塩基	93.35%	95.58%
Q30以上のリード2の塩基	92.21%	94.24%
リード1のエラー率	0.08%	0.09%
リード2のエラー率	0.28%	0.16%
平均されたサンプル数	10	80
サンプルあたりの推測される細胞数	3,880	3,882

単一のフローセルランからのメトリクスは、レーン数が異なる複数のフローセル全体を平均化したものです。すべてのランは出力に関して公表されている仕様を満たしました。レーンあたりの出力は、NovaSeq 6000 S4フローセルとNovaSeq X 10Bフローセルの間で同等ではありません。

まとめ

NovaSeq XおよびNovaSeq X Plusシステムは、ブレイクスルーケミストリー、光学系、インフォマティクス、および操作の簡便性を特徴としており、ハイスループットシーケンスの経済性を変革します。NovaSeq Xシリーズは、ユーザーがイルミナに期待する高品質のデータを提供しながら、驚異的なスループットを実現します。XLEAP-SBSケミストリーにより、データ品質を犠牲にすることなく、シーケンスランタイムと出力を大幅に改善できます。全ゲノムシーケンス、全エクソームシーケンス、全トランスクリプトームシーケンス、メチル化シーケンス、シングルセルマルチオミクスなど、NovaSeq 6000システムで一般的に実行される主要なメソッドからのデータを、NovaSeq X Plusシステムを使用して生成されたデータと直接比較しました。これらの結果は、NovaSeq Xシリーズの性能がNovaSeq 6000システムの性能を満たすか、それを上回り、より多くのデータ集約型アプリケーションに対応することを示しています。

詳細はこちら

NovaSeq XおよびNovaSeq X Plusシステム: jp.illumina.com/systems/sequencing-platforms/novaseq-x-plus.html

このノートで参照されたデータセット: basespace.illumina.com/datacentral

参考文献

1. National Institute of Standards and Technology. Genome in a Bottle. nist.gov/programs-projects/genome-bottle. Accessed July 27, 2023.
2. Genome Reference Consortium. Human Genome Overview. NCBI website. ncbi.nlm.nih.gov/grc/human. Accessed July 27, 2023.
3. Illumina. Best practices for read trimming for Illumina Stranded mRNA and Total RNA workflows. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-stranded-rna-t-overhang-tech-note-470-2020-010/illumina-stranded-rna-t-overhang-tech-note-470-2020-010.pdf. Published 2020. Accessed August 1, 2023.
4. Wehrkamp-Richter S. Illumina.BaseSpace™ Sequence Hub now supports whole genome bisulfite sequencing (WGBS) data with Zymo-Seq Library Kit running on NovaSeq X Series. developer.illumina.com/news-updates/whole-genome-bisulfite-sequencing-zymo-seq-data-now-available-on-novaseqtm-x-series. Published June 6, 2023. Accessed August 1, 2023.

イルミナ株式会社

〒108-0014 東京都港区芝 5-36-7 三田ベルジュビル 22階
Tel (03) 4578-2800 Fax (03) 4578-2810
jp.illumina.com

 www.facebook.com/illuminakk

販売店

本製品の使用目的は研究に限定されます。診断での使用はできません。 販売条件: jp.illumina.com/tc

© 2023 Illumina, Inc. All rights reserved.
すべての商標および登録商標は、Illumina, Inc. または各所有者に帰属します。
商標および登録商標の詳細は jp.illumina.com/company/legal.html をご覧ください。
予告なしに仕様および希望販売価格を変更する場合があります。

